

CLAIMS:

1. A method of regulating admittance of a number of TCP connection requests to a plurality of servers that support to a plurality of applications and a plurality of customers, each of the customers having electronic business activity hosted by the plurality of servers on the basis of a service level agreement that specifies a minimum TCP connection rate that the customer is guaranteed, the method comprising the steps of:

receiving incoming workload for the plurality of customers for the plurality of applications into a common buffer, the incoming workload comprising IP packets associated with existing TCP connections and TCP connection requests; and then

controlling flow of the incoming workload from the common buffer to the plurality of servers for each application and each customer so as to provide at least the minimum TCP connection rate for each customer and each application by regulating flow of the TCP connection request packets to the plurality of servers.

2. The method according to claim 1, wherein all of the incoming workload is received into the common buffer and the TCP connection request packets are processed on a first in, first out basis.

3. The method according to claim 1, wherein the method does not directly control outbound traffic from the plurality of servers.

4. The method according to claim 1, wherein the controlling step is performed by detecting whether an IP packet is a TCP connection request packet, and immediately admitting all other packets.

5. The method according to claim 4, wherein the detecting step is

performed by determining whether a SYN bit of the IP packet is ON or OFF.

6. The method according to claim 4, wherein IP packets associated with TCP connection requests are admitted if necessary to meet the minimum TCP connection rate for the customer associated with the TCP connection request.

7. The method according to claim 6, wherein IP packets associated with a TCP connection request of one of the customers are also admitted if doing so does not prevent meeting the minimum TCP connection rate for another of the plurality of customers.

8. The method according to claim 7, further comprising the step of dropping TCP connection request packets that are not admitted during the controlling step.

9. The method according to claim 7, further comprising the step of returning TCP connection request packets that are not admitted during the controlling step, the TCP connection request packets not admitted during the controlling step being returned with a reset code bit ON.

10. The method according to claim 1, further comprising the step of calculating whether TCP connection request packets of one or more of the customers can be admitted for one or more of the applications when the one or more customers already exceed the minimum TCP connection rate for the one or more applications.

11. The method according to claim 1, further comprising the step of communicating with an external means to receive revisions to one or more of the service level agreements.

12. The method according to claim 1, further comprising the step of communicating with an external means to send collected statistics data to the external means.

13. The method according to claim 1, wherein the controlling step is performed with a real-time admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the plurality of servers on a per-packet basis.

14. The method according to claim 1, wherein the controlling step is performed with a target-rate-based admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the plurality of servers on a per-unit-time basis.

15. A method of controlling and guaranteeing a service level agreement based on a number of TCP connection requests supported to a plurality of applications and a plurality of customers having electronic business activity hosted by a server farm, each of the customers having a service level agreement that specifies a minimum TCP connection rate that the customer is guaranteed, the method comprising the steps of:

receiving all incoming workload for the plurality of customers for the plurality of applications into a common buffer, the incoming workload comprising IP packets associated with existing TCP connections and TCP connection requests;

taking TCP connection request packets from the common buffer on a first in, first out basis; and

controlling flow of the IP packets from the common buffer to the server farm for each application and each customer so as to provide at least the minimum TCP connection rate for each customer and each application by regulating flow of the TCP connection request packets to the server farm, the controlling step

comprising:

detecting whether an IP packet is a TCP connection request packet or associated with an existing TCP connection;

immediately admitting IP packets associated with existing TCP connections;

admitting TCP connection request packets if necessary to meet the minimum TCP connection rate for the customer associated with the TCP connection request; and

admitting a TCP connection request packet of one of the customers if doing so does not prevent meeting the minimum TCP connection rate for another of the plurality of customers.

16. The method according to claim 15, wherein the method does not directly control outbound traffic from the serve farm.

17. The method according to claim 15, wherein the detecting step is performed by determining whether a SYN bit of the IP packet is ON or OFF.

18. The method according to claim 15, further comprising the step of dropping IP packets that are not admitted during the controlling step.

19. The method according to claim 15, further comprising the step of returning TCP connection request packets that are not admitted during the controlling step, the IP packets not admitted during the controlling step being returned with a reset code bit ON.

20. The method according to claim 15, further comprising the step of calculating whether a TCP connection request packet of one or more of the customers can be admitted for one or more of the applications when the one or more customers

already exceed the minimum TCP connection rate for the one or more applications.

21. The method according to claim 15, further comprising the step of communicating with an external means to receive revisions to one or more of the service level agreements, and to send collected statistics data to the external means.

22. The method according to claim 15, wherein the controlling step is performed with a real-time admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the server farm on a per-packet basis.

23. The method according to claim 15, wherein the controlling step is performed with a target-rate-based admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the server farm on a per-unit-time basis.

24. A system for regulating admittance of a number of TCP connection requests to a plurality of servers that support to a plurality of applications and a plurality of customers, each of the customers having electronic business activity hosted by the plurality of servers on the basis of a service level agreement that specifies a minimum TCP connection rate that the customer is guaranteed, the system comprising:

a common buffer for receiving incoming workload for the plurality of customers for the plurality of applications, the incoming workload comprising IP packets associated with existing TCP connections and TCP connection requests; and

means for controlling flow of the incoming workload from the common buffer to the plurality of servers for each application and each customer so as to provide at least the minimum TCP connection rate for each customer and each application by regulating flow of the TCP connection request packets to the plurality of servers.

25. The system according to claim 24, wherein the common buffer is configured to receive all of the incoming workload, and the controlling means is configured to process the TCP connection request packets on a first in, first out basis.

26. The system according to claim 24, wherein the system is not configured to directly control outbound traffic from the plurality of servers.

27. The system according to claim 24, wherein the controlling means is configured to detect whether an IP packet is associated with an existing TCP connection or a TCP connection request, and to immediately admit IP packets associated with existing TCP connections.

28. The system according to claim 27, wherein the controlling means is configured to determine whether a SYN bit of an IP packet is ON or OFF.

29. The system according to claim 4, wherein the controlling means is configured to admit TCP connection request packets if necessary to meet the minimum TCP connection rate for the customer associated with the TCP connection request.

30. The system according to claim 29, wherein the controlling means is configured to admit a TCP connection request packet of one of the customers if doing so does not prevent meeting the minimum TCP connection rate for another of the plurality of customers.

31. The system according to claim 30, further comprising means for dropping TCP connection request packets that are not admitted by the controlling means.

32. The system according to claim 30, further comprising means for returning TCP connection request packets that are not admitted with a reset code bit ON.

33. The system according to claim 24, further comprising means for calculating whether TCP connection request packets of one or more of the customers can be admitted for one or more of the applications when the one or more customers already exceed the minimum TCP connection rate for the one or more applications.

34. The system according to claim 24, further comprising means for communicating with an external means to receive revisions to one or more of the service level agreements.

35. The system according to claim 24, further comprising means for communicating with an external means to send collected statistics data to the external means.

36. The system according to claim 24, wherein the controlling means operates with a real-time admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the plurality of servers on a per-packet basis.

37. The system according to claim 24, wherein the controlling means operates with a target-rate-based admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the plurality of servers on a per-unit-time basis.

38. A system of controlling and guaranteeing a service level agreement based on a number of TCP connection requests supported to a plurality of applications

and a plurality of customers having electronic business activity hosted by a server farm, each of the customers having a service level agreement that specifies a minimum TCP connection rate that the customer is guaranteed, the system comprising:

a common buffer into which is received all incoming workload for the plurality of customers for the plurality of applications, the incoming workload comprising IP packets associated with existing TCP connections and TCP connection requests;

a gatekeeper that takes the IP packets from the common buffer on a first in, first out basis and controls flow of the IP packets from the common buffer to the server farm for each application and each customer so as to provide at least the minimum TCP connection rate for each customer and each application by regulating flow of the IP packets to the server farm, the gatekeeper being operable to detect whether an IP packet is a TCP connection request packet or associated with an existing TCP connection, immediately admit IP packets associated with existing TCP connections, admit a TCP connection request packet if necessary to meet the minimum TCP connection rate for the customer associated with the TCP connection request, and admit a TCP connection request packet of one of the customers if doing so does not prevent meeting the minimum TCP connection rate for another of the plurality of customers.

39. The system according to claim 38, wherein the system does not directly control outbound traffic from the serve farm.

40. The system according to claim 38, wherein the gatekeeper is configured to determine whether a SYN bit of the IP packet is ON or OFF.

41. The system according to claim 38, further comprising a guide means for dropping TCP connection request packets that are not admitted by the

gatekeeper

42. The system according to claim 38, further comprising a guide means for returning TCP connection request packets that are not admitted with a reset code bit ON.

43. The system according to claim 38, further comprising an assistant means for calculating whether TCP connection request packets of one or more of the customers can be admitted for one or more of the applications when the one or more customers already exceed the minimum TCP connection rate for the one or more applications.

44. The system according to claim 38, further comprising a manager means for communicating with an external means to receive revisions to one or more of the service level agreements, and to send collected statistics data to the external means.

45. The system according to claim 38, wherein the gatekeeper operates with a real-time admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the server farm on a per-packet basis.

46. The system according to claim 38, wherein the gatekeeper operates with a target-rate-based admittance/rejection algorithm wherein the flow of the TCP connection request packets is regulated to the server farm on a per-unit-time basis.